Eur. J. Biochem. 251, 549–557 (1998)

Review

Human aldehyde dehydrogenase gene family

Akira YOSHIDA[1], Andrey RZHETSKY[2], Lily C. HSU[1] and Cheng CHANG[1]

[1] Beckman Research Institute of the City of Hope, Duarte CA, USA
[2] Columbia Genome Center, Columbia University, New York NY, USA

(Received 3 July/17 November 1997) – EJB 97 0937/0

Twelve aldehyde dehydrogenase (ALDH) genes have been identified in humans. These genes, located on different chromosomes, encode a group of enzymes which oxidizes varieties of aliphatic and aromatic aldehydes. Metabolic disorders and clinical problems associated with mutations of *ALDH1*, *ALDH2*, *ALDH4*, *ALDH10* and succinic semialdehyde (*SSDH*) genes have been emerged. Comparison of the human ALDHs indicates a wide range of divergency (>80–<15% identity at the protein sequence level) among them. However, several protein regions, some of which are implicated in functional activities, are conserved in the family members.

The phylogenic tree constructed of 56 ALDH sequences of humans, animals, fungi, protozoa and eubacteria, suggests that the present-day human *ALDH* genes were derived from four ancestral genes that existed prior to the divergence of Eubacteria and Eukaryotes. The neighbor-joining tree derived from 12 human ALDHs and antiquitin indicates that diversification within the *ALDH1/2/5/6* gene cluster occurred during the Neoproterozoic period (about 800 million years ago). Duplication in the *ALDH 3/10/7/8* gene cluster occurred in Phanerozoic period (about 300 million years ago). Separations of *ALDH3/ALDH10* and that of *ALDH7/ALDH8* had occurred during the period of appearance and radiation of mammalian species.

Keywords: gene family; genomic organization; genetic disease; genetic variant; detoxification; evolution; phylogenetic tree.

Aldehyde dehydrogenases [aldehyde: NAD(P)$^+$ oxidoreductase] are a group of enzymes catalyzing the conversion of aldehydes to the corresponding acids by means of an NAD(P)$^+$-dependent virtually irreversible reaction. ALDHs are widely distributed from bacteria to humans.

Mammalian ALDH activity was first observed in ox liver nearly 50 years ago [1] and thereafter several types of ALDH were distinguished based on their physico-chemical characteristics, enzymological properties, subcellular localization, and tissue distribution [2–4]. Two *ALDH* genes were cloned and characterized in 1985 [5]. At the present time, ten non-allelic genes have been identified in the human ALDH family. In addition, partial cDNAs for two distantly related ALDHs, i.e. succinic semialdehyde dehydrogenase (SSDH) and methylmalonate semialdehyde dehydrogenase (MMSDH) were also reported [6, 7]. Most, if not all, corresponding members of the ALDH family probably exist in other mammals. Protein sequences, genes and/or cDNAs for more than 50 animals, fungi, and bacterial ALDHs have been reported.

Correspondence to A. Yoshida, Dept. of Biochemical Genetics, Beckman Research Institute of the City of Hope, Duarte, CA 91010 USA

Abbreviations. ALDH, aldehyde dehydrogenase; SSDH, succinic semialdehyde dehydrogenase; MMSDH, methylmalonate semialdehyde dehydrogenase; γABDH, 4-aminobutyraldehyde dehydrogenase; FALDH, fatty aldehyde dehydrogenase; 4-Abu, 4-aminobutyric acid; nt, nucleotide.

Note. This Review will be reprinted in *EJB Reviews 1998* which will be available in April 1999.

This paper reviews the functional and structural diversity and evolution of the human ALDH gene family.

There is no uniform nomenclature system for human and animal *ALDH* genes and enzymes. Therefore, commonly used abbreviated human gene symbols (GBD symbols) are used for genes (in italic) and enzymes (in non-italic) in the present review. GenBank identification numbers are also given.

Members of ALDH families

Twelve known human *ALDH* genes and corresponding enzymes are listed in Table 1. These genes consist of 10–13 coding exons and span 11–40 kbp in various chromosomes. *ALDH5* does not contain introns in the coding region. Genomic organizations of *ALDH4*, *SSDH* and *MMSDH* are not yet known.

Proteins (enzyme subunits) encoded by these genes consist of about 500 amino acid residues. Catalytically active forms of the enzymes are homodimers (ALDH3, ALDH4), homotetramers (ALDH1, ALDH2, ALDH9, MMSDD) or unknown.

Tissue distribution and subcellular localization of individual ALDHs are also shown in Table 1.

More recently, an additional human cDNA, tentatively designated as ALDH11 cDNA, was cloned. This cDNA encoding 499 amino acid residues, is strongly expressed in testis, and its deduced amino acid sequence is highly similar (72%) to human ALDH1 (Hsu et al. unpublished observation). Since the sequence and properties of ALDH11 have not yet been published, it is not included in Table 1.

Table 1. Human ALDH family. Tissue (subcellular) distribution indicates the major tissue expressing a given ALDH member at a high level; (cyt) indicates cytoplasm, (mit) mitochondria, (micro) microsomes. The subcellular distribution of ALDH5, ALDH7 and ALDH8 is based on the presence of an NH_2-terminal leader sequence or COOH-terminal anchor sequence, not on the separation of subcellular components. Major substrate indicates aldehydes which are effectively oxidized by a given isozyme. Coding sequence indicates the number of deduced amino acid residues, including the chain initiator Met. Subunit size indicates the number of amino acid residues of subunits which constitute catalytically active enzymes, excluding the chain initiation Met and the NH_2-terminal signal peptides in the case of ALDH2, ALDH4 and ALDH5. In the final column, references are given to papers which originally described full-length cDNA and genomic organization. For SSDH and MMSDH, papers reporting partial cDNAs are cited. References for tissue and subcellular distribution, major substrates and chromosomal locations are given in the literature cited here.

Gene (GDB symbol)	Enzyme (abbreviated symbol)	Tissue (subcellular distribution)	Major substrate	Coding sequence (amino acid)	Subunit size (amino acid)	Chromosomal location	Reference (GenBank I.D.)
ALDH1	ALDH1	liver, stomach etc. (cyt)	retinal	501	500	9q21	[8] (J04748)
ALDH2	ALDH2	liver, stomach etc. (mit)	acetaldehyde	517	500	12q24	[9] (M20444)
ALDH3	ALDH3	stomach, lung etc. (cyt)	fatty and aromatic aldehydes	453	452	17p11.2	[10, 11] (M77491)
ALDH4	ALDH4	liver, kidney (mit)	glutamate γ-semialdehyde	563	539?	1	[12] (U24266)
ALDH5	ALDH5	testis, liver (mit)	propionaldehyde	517	500	9p13	[13] (M63967)
ALDH6	ALDH6	salivary gland, stomach, kidney (cyt)	aliphatic aldehyde, retinal	512	511	15q26	[14] (U07919)
ALDH7	ALDH7	kidney, lung (micro)	aliphatic and aromatic aldehydes	468	467	11q13	[15, 16] (U10867)
ALDH8	ALDH8	parotid (micro)	unknown	451	450	11q13	[16] (U37519)
ALDH9	γABDH	liver, kidney, muscle (cyt)	amine aldehyde	493	492	1q22–24	[17] (U34252)
ALDH10	FALDH	liver, heart, muscle (micro)	fatty and aromatic aldehydes	485	484	17p11	[18, 19] (U46689)
SSDH	SSDH	brain, liver, heart (mit)	succinic semialdehyde	?	488	6	[6, 20] (L34821)
MMSDH	MMSDH	kidney, liver, heart (mit)	methylmalonate semialdehyde	535*	503*	?	[7] (M93405)

* Deduced from the homologous rat cDNA and enzyme.

Human antiquitin cDNA (GenBank S74728) [21] is distantly related (similarity 15–25%) to human ALDHs. Human antiquitin is similar (about 60%) to a hypothetical ALDH-like protein (Swiss-Prot P46562; GenBank U13070) identified in *Caenorhabditis elegans*. It is not yet known whether or not human antiquitin has ALDH activity.

Functional diversity

ALDHs exhibit a rather broad substrate specificity and many of them can oxidize varieties of aliphatic and aromatic aldehydes. For most ALDHs, NAD is a better co-enzyme than NADP and the enzymes also have esterase activity.

ALDHs have been considered as general detoxifying enzymes which eliminate toxic biogenic and xenobiotic aldehydes [22–24]. More recently, specific biological roles of some ALDHs have emerged.

ALDH1. ALDH1 is a cytosolic enzyme ubiquitously distributed in various tissues including brain and red blood cells. The enzyme has a high activity for oxidation of both all-*trans*- and 9-*cis*-retinal ($K_m < 0.1$ μM for all-*trans*-retinal at pH 7.5) [25], and it may play a vital role in the formation of retinoic acid which is a potent modulator for gene expression and tissue differentiation. *ALDH1* gene is not expressed in genital tissues of patients associated with androgen receptor-negative testicular feminization, although the gene is expressed in other tissues [26, 27]. It was suggested that activation of *ALDH1* gene is mediated by androgen receptor and generation of retinoic acid by ALDH1 is required for testicular development.

The recently identified human ALDH11 described in the preceding section, may also participate in the regulation of gene expression and tissue differentiation mediated by retinoic acid, since the corresponding mouse and rat ALDHs, which are highly similar (> 95%) to ALDH11, have a high activity for oxidation of retinal [28, 29].

Several ALDH1 variants, associated with various degrees of enzyme deficiency in the liver and red blood cells, have been reported [30–32]. Nucleotide changes of these variants have not been determined. Except for possible alcohol sensitivity, physiological problems were not observed in the variant subjects.

ALDH1 exhibits a high activity for oxidation of aldophosphamide and plays a role for detoxification of widely used anticancer drugs, oxazaphosphorines [33, 34]. It was demonstrated that the acquired drug resistance was associated with the transcriptional activation of *ALDH1* expression in the cells [35].

Porcine cytosolic ALDH, which corresponds to human ALDH1, has a high activity for oxidation of 11-hydroxy-thromboxane B2 and may participate in the thromboxane metabolism [36].

ALDH1 is a major soluble constituent of eye lens and may play a role in detoxification of peroxidic aldehydes produced by ultraviolet light absorption [37].

ALDH2. ALDH2 is a mitochondrial enzyme strongly expressed in various tissues with the highest level in the liver. ALDH2 exhibits a high activity for oxidation of acetaldehyde ($K_m < 5$ μM at pH 7.5), and plays a major role in acetaldehyde detoxification. The alcohol sensitivity (i.e. facial flushing, elevation of skin temperature and increase in pulse rate, etc.) in Orientals is associated with the genetic deficiency of ALDH2 caused by a point mutation G→A transition in exon 12, Glu→Lys at 487 position from NH_2-terminal Ser of the matured subunit (review in [38]). Other type of *ALDH2* mutation was found in American Indians associated with the enzyme deficiency [39].

Since no adverse developmental or physiological problems are observed in homozygous variant subjects associated with vir-

tually null ALDH2 activity, ALDH2 is probably not essential for survival.

ALDH3. ALDH3 oxidizes aromatic aldehydes and medium-chain aliphatic aldehydes (fatty aldehydes). The enzyme is a cytosolic enzyme strongly expressed in the stomach and lung, but at a low level (or undetectable) in the normal liver. ALDH3 is strongly expressed in about 70% of poorly differentiated and 30% of well differentiated human hepatocellular carcinomas [40]. Etiological relationships of hepatoma and ALDH3 expression are not clear.

Stable ALDH3 expression was observed in the carcinogen-induced rat hepatoma, and transient expression in cultured hepatocytes treated with aromatic hydrocarbon xenobiotics (review in [41]).

The oxazaphosphorine resistance was substantially elevated in hamster cell lines transfected with rat or human ALDH3 cDNA constructs [42]. It was observed that acquired oxazaphosphorine resistance was accompanied with the enhanced expression of ALDH3-like enzyme in human breast and paroid gland carcinoma cell lines [43–45].

However, activity toward aldophosphamide with purified stomach ALDH3 and the ALDH3-like enzyme prepared from the resistant cells is too weak to account for the drug resistance [42–45]. Thus, it was suggested that aldophosphamide oxidation activity of ALDH3 could be more severely diminished than benzaldehyde oxidation activity during purification, or that ALDH3 may be extremely sensitive to inhibition by acrolein, which would likely react with many nucleophiles and be eliminated *in vivo* [42].

ALDH3 exhibits 3–5 components distinguishable in isoelectric focussing. It was suggested that two subunits, encoded by two non-allelic genes, were involved in formation of homomeric and heteromeric oligomers [46]. Such a possibility was excluded by genomic analysis [10]. The *ALDH3* gene has multiple transcription initiation sites and three mRNAs, with same coding sequence but with different 5'-untranslated sequence lengths, are produced from the transcripts [11]. *Escherichia coli* transfected with single cloned human ALDH3 cDNA produces multiple ALDH3 components. Thus, ALDH3 multiplicity must be due to post-translational protein modification.

A common variant *ALDH3* allele (variant allele frequency about 0.25, C→G transversion at nt 985, Pro→Ala at protein position 329) exists in both Caucasians and Orientals [47]. According to the recent X-ray crystallographic study of rat ALDH3, Pro329 exists at the end of short β-9, and a Pro→Ala substitution could cause structural and functional changes [48]. However, no physiological problems were reported in the variant subjects.

The existence of ALDH3 variant(s), distinguished by electrophoresis or isoelectricfocusing, was reported in Caucasians and Orientals [46, 49]. The mutation site(s) of the variant(s) has not been determined, and it is not clear whether or not the variant(s) is identical to the common variant described above.

ALDH3 is a major constituent of cornea in humans and other animals [37, 50] and, like ALDH1, it may play a role in detoxification of peroxidic aldehydes.

ALDH4. ALDH4 is a mitochondrial enzyme with a high activity for oxidation of γ-semialdehydes such as glutamic γ-semialdehyde (a hydration product of pyrroline 5-carboxylate) [12, 51]. Genetic deficiency of ALDH4 disturbs proline degradation and 4-Abu formation, causing type II hyperprolinemia, associated with elevation of plasma proline level, mental retardation and convulsion [52, 53]. The mutation site of the defective gene(s) has not been determined.

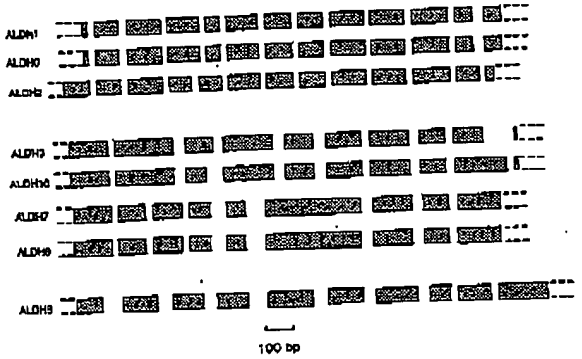


Fig. 1. Comparison of genomic organization of eight human *ALDH* genes. Exons containing the translated regions (shadowed) are shown. *ALDH5* does not contain introns in its coding exon; the genomic organization of other human *ALDH* genes is unknown.

ALDH9 (γABDH). γABDH is a cytosolic enzyme with a high activity for oxidation of 4-aminobutyraldehyde [54]. The *ALDH9* gene is expressed at a low level in adult brain, but it is strongly expressed in the early stage of embryonic brain (gestational age of < 12 weeks) [17, 54]. 4-Abu may be produced from putrescine rather than from glutamate, by diamine oxidase and γABDH in the mammalian embryonic brain, as has been observed in the avian embryonic brain [55].

The *ALDH9* locus is polymorphic, i.e. C or T at nt 327, and G or C at nt 344. The most common (>50% in Caucasians) haplotype is C at nt 327 and G at nt 344, and haplotypes of T327/G344 and C327/C344 are less common (about 20% each) [17]. The transition C→T at nt 327 is silent, but the transversion G→C at nt 344 should induce the amino acid substitution Cys →Ser at protein position 115 and could alter enzyme properties.

The notion for the existence of tissue-specific type of γABDHs (liver enzyme with Ser, brain enzyme with Cys) [56] is untenable. The difference is due to the existence of polymorphism at the single *ALDH9* locus.

ALDH10 (FALDH). FALDH is a microsomal enzyme with a high activity for oxidation of medium-chain aliphatic aldehydes (fatty aldehydes). The genetic deficiency of FALDH disturbs the metabolism of membrane lipid causing Sjögren-Larsson syndrome, an inherited disorder characterized by ichthyosis, neurological problems and oligophrenia [57]. Five distinctive types of mutations were found in the patients examined [47, 57]. As described above, ALDH3 also has a high activity for oxidation of fatty aldehydes. Although organization and structure of *ALDH3* and *ALDH10* are highly similar, *ALDH10* has a longer exon 9 and encodes the COOH-terminal transmembrane domain producing microsomal enzyme, while *ALDH3* lacks the corresponding sequence in exon 9, producing a cytosolic enzyme (Fig. 1). Thus, ALDH3 probably cannot supplement the role of FALDH in the synthesis of membrane lipid.

SSDH. SSDH, which is strongly expressed in brain, has a high activity for oxidation of succinic semialdehyde [20]. Genetic deficiency of SSDH disturbs the oxidation of succinic semialdehyde which is produced from 4-Abu by 4-Abu–Glu transaminase, causing accumulation of succinic semialdehyde and psychomotor retardation [58]. The mutation site of the gene has not been determined.

In addition to SSDH, adult brain contains ALDH1, ALDH2 and ALDH5. γABDH is also expressed at a low level in adult

Table 2. Comparison of amino acid sequences in the ALDH family. Maximum identity and diasimilarity were compared by the Clustal V method (Megalign program, DNASTAR, Inc., Madison WI).

Percent Similarity

Percent Divergence

	1	2	3	4	5	6	7	8	9	10	11	12		
1		69.9	67.9	64.3	35.7	32.2	25.8	23.4	20.6	19.0	19.7	16.8	1	ALDH1.PRO
2	29.9		63.5	61.7	32.3	32.2	24.9	22.3	20.0	20.7	21.2	19.9	2	ALDH6.PRO
3	32.3	35.7		72.1	38.3	31.7	24.2	22.1	18.8	18.6	18.5	19.0	3	ALDH2.PRO
4	35.6	37.8	27.1		35.9	31.7	24.9	22.1	18.8	20.1	19.7	18.6	4	ALDH5.PRO
5	59.2	63.1	57.9	59.9		31.7	24.2	23.4	21.0	20.1	19.5	15.8	5	ALDH9.PRO
6	60.9	60.6	62.5	61.9	61.7		22.4	22.1	18.8	21.6	19.3	15.1	6	SSDH.PRO
7	69.4	69.4	69.8	70.1	72.2	70.4		18.8	17.9	17.4	17.2	15.3	7	MMSDH.PRO
8	70.8	72.0	72.0	71.5	71.5	72.9	77.5		65.1	52.8	48.8	15.0	8	ALDH3.PRO
9	72.7	73.9	74.6	74.6	74.2	73.8	77.5	34.2		52.4	47.6	12.0	9	ALDH10.PRO
10	74.9	73.3	75.4	74.2	74.5	73.4	79.1	46.7	46.7		81.3	14.1	10	ALDH7.PRO
11	74.0	72.1	74.9	74.0	76.4	74.6	79.8	50.9	51.6	18.0		13.7	11	ALDH8.PRO
12	73.0	73.2	75.3	75.7	75.1	75.8	78.4	78.9	80.7	79.0	78.5		12	ALDH4.PRO
	1	2	3	4	5	6	7	8	9	10	11	12		

brain. ALDH1, ALDH2 and ALDH3 can oxidize dopaldehyde (3, 4-dihydroxyphenyl acetaldehyde) [59], but their roles in dopamine homeostasis are not clear.

An ALDH isozyme, which could be attributed to *ALDH5* gene product, exhibited a high activity for oxidation of short-chain aliphatic aldehydes [60].

MMSDH is the only CoA-dependent dehydrogenase in the ALDH family. The enzyme catalyzes oxidative decarboxylation of malonate semialdehyde and methyl malonate semialdehyde to acetyl-CoA and propionyl-CoA, and may be involved in the catabolism of β-alanine, valine and thymine [61].

ALDH 6 has activity for retinal oxidation (Hsu et al., unpublished observation), and may be involved in retinoic acid formation. Biological substrates of ALDH7 and ALDH8 are unknown.

Structural similarity

A comparison of the amino acid sequences of human ALDHs indicates a wide range of divergency among the members (Table 2). The maximum identity, adjusting for gaps, ranges from >80% to <15%. ALDH4, which is one of the classical human ALDH members, is not highly similar (<20%) to any of the other members.

A comparison of their intron-exon organizations revealed that the gene family can be divided into group 1/6/2, group 3/10, group 7/8, and a unique *ALDH9* which seems to have branched off from group 3/10. Intron-exon junction positions and sizes of coding sequences in each exon coincide within each group (Fig. 1).

Organization of *ALDH7* and *ALDH8* is similar to that of *ALDH3* and *ALDH10*, except for the difference of intron positions in exons 2–6 between the two groups. The *ALDH5* gene does not have introns within the coding region, and genomic organization of *ALDH4*, *SSDH* and *MMSDH* are not known, so these genes cannot be included in Fig. 1.

Detailed comparison of the amino acid sequences and intron-exon organization indicate the existence of a remarkable similarity among all human ALDH family members (Fig. 2).

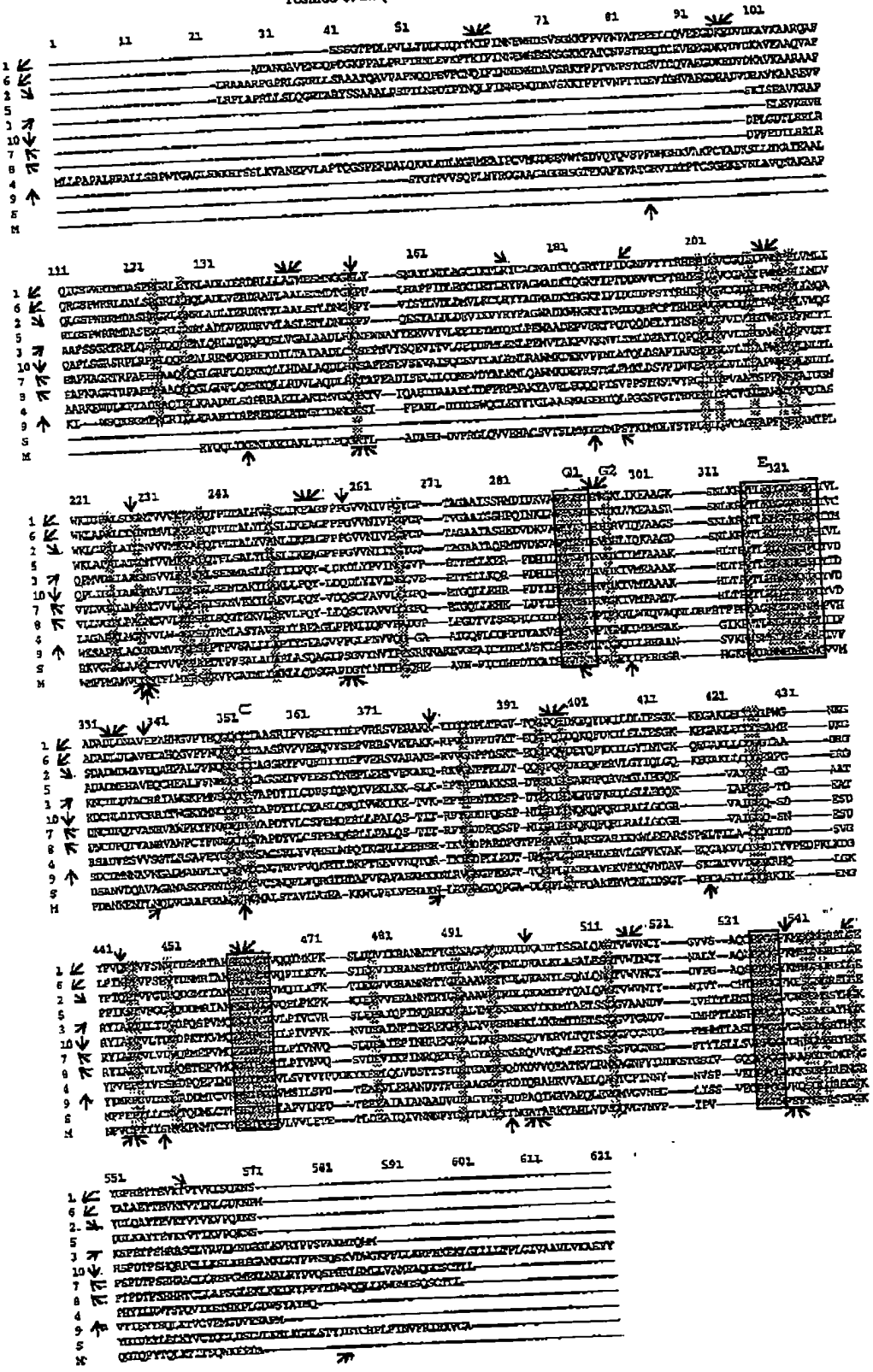
Peptide regions shown in Fig. 2 in open boxes and amino acid residues in shadowed boxes are conserved in all or most of ALDHs. Functional roles of some of these conserved residues have been identified by site-directed mutagenesis and X-ray crystallographic studies. The conserved Glu (marked E) is important for catalytic activity. Substitution of this Glu by other amino acid residue did not affect K_m values for NAD or propionaldehyde, of human ALDH2, but grossly reduced the catalytic activity [62]. It is suggested that this Glu residue is not directly involved in substrate binding but it functions as a general base necessary for the activation. Three-dimensional analysis of bovine ALDH2 supports this notion [63].

Site-directed mutagenesis of rat ALDH2 revealed that the conserved Cys (marked C) is the single active-site nucleophil which forms a covalent bond with the substrate. Substitution of this Cys by a poor nucleophil caused severe loss of a catalytic activity [64]. Three-dimensional analysis of bovine ALDH2 [63] and rat ALDH3 [65] supported this conclusion.

Two Gly residues (marked G1 and G2) are found to participate in NAD binding, based on three-dimensional analysis of rat ALDH3 [65]. The conserved Lys residue (normalized amino

Fig. 2. Alignment of 12 amino acid sequences of the human ALDH family. Sequences denoted 1, 6, 2, 5, 3, 10, 7, 8, 4, 9, S and M correspond to ALDH1, ALDH6, ALDH2, ALDH5, ALDH3, ALDH10, ALDH7, ALDH8, ALDH4, ALDH9, SSDH and MMSDH, respectively. The intron insertion positions are indicated by arrowheads. ↙ for 1 and 6, ↘ for 2, shown above the alignment, and ↙ for 3, ↓ for 10, ↘ for 7 and 8, and ↓ for 9, shown below the alignment. Open boxes emphasize conserved regions with conserved amino acid residues in shaded boxes. Note that only partial coding sequences are currently available for human *SSDH* and *MMSDH*. Intron-exon organization of *ALDH4*, *SSDH* and *MMSDH* is unknown, and *ALDH5* does not have introns in its coding region.

Yoshida et al. (*Eur. J. Biochem. 251*)



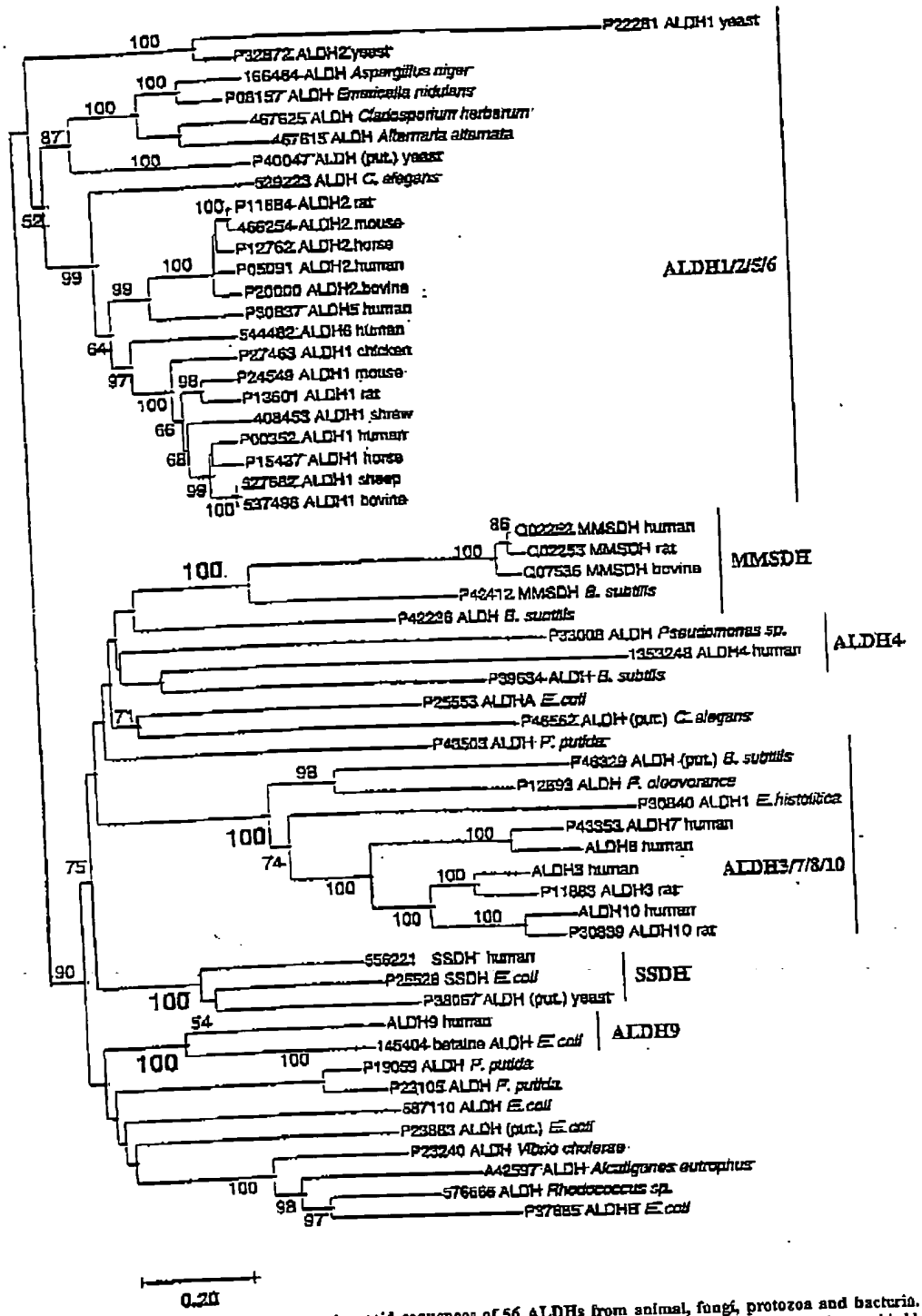


Fig. 3. A neighbor-joining tree derived from amino acid sequences of 56 ALDHs from animal, fungi, protozoa and bacteria. The tree was constructed by the formula of Saitou and Nei [73], and visualized with computer programs NJBootW and TreeView (kindly provided by K. Tamura). The distances between sequences were computed using Poisson correction for multiple hits [74] and are expressed in terms of the amino acid substitutions/site. Bootstrap *P*-values were computed from 1000 bootstrap resamplings [75]. Deposition of each protein includes either Swiss Prot or GenBank accession number, protein name and species name. In order to avoid over complexity, the unique plant betaine aldehyde dehydrogenase group is not included in the tree.

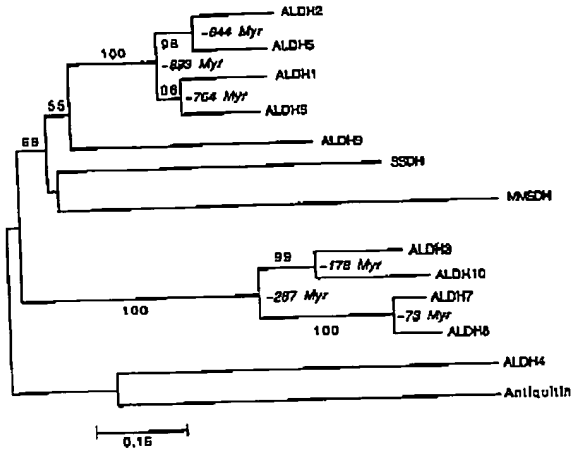


Fig. 4. A neighbor-joining tree derived from 12 human ALDHs and antiquitin. All sites containing insertions or deletions were excluded from data analysis. Bootstrap values above 50% are indicated. The times of divergence were estimated by linearized tree algorithm I J [76] based on the known ALDH protein sequences from rodents, primates and artyodactyls, and on the divergence time 104 million years (Myr) ago for Rodentia/Artiodactyla bifurcation [77].

acid position 236 in Fig. 2) was also implicated in NAD binding by a site-directed mutagenesis study of human ALDH2 [66, 67].

Evolution of ALDH

A general neighbor-joining tree was derived from 56 ALDH sequences of humans, animals, fungi, protozoa and bacteria (Fig. 3). A human neighbor-joining tree was derived from 12 human ALDHs and antiquitin (Fig. 4).

There is inconsistency in the branch point of human SSDH and MMSDH between the two neighbor-joining trees, i.e. they are clustered in the human tree (Fig. 4) while they belong to separate clusters in the general tree (Fig. 3). Since bootstrap *P*-values of these clusters are low (<20%) in both trees, the origin of these two genes is not certain. The branch point(s) of *ALDH4* and antiquitin is also not certain.

Except for the uncertainties described above, all cluster branch points have very high bootstrap *P*-values, indicating a high degree of validity in the trees. The trees indicate that diversification within the ALDH 1/2/5/6 cluster was likely to occur much earlier than within the ALDH 3/10/7/8 cluster (Figs 3 and 4). In the reconstruction, diversification within ALDH 1/2/5/6 cluster happened during the Neoprotozoic period ((893 million years ago), long before the Ediacaran radiation of metazoan [68]. According to the current paleontological paradigm, the earth during this period was dominated by unicellular organisms, and multicellular organisms, mostly algae, were scarce and primitive [69].

In contrast, multiplication in the ALDH 3/10/7/8 cluster seemed to arise much later (≈287 million years ago) in the Phanerozoic period, when diverse vertebrate and invertebrate animals were abundant. The latest two duplications ALDH3/ALDH10 and ALDH7/ALDH8 took place about 178 and 73 million years ago, i.e. corresponding to the periods of early appearance and radiation of mammalian species.

The separation of two major groups (i.e. ALDH1/2/5/6/9 and ALDH3/10/7/8) had occurred more than two billion years ago [70], before the separation of the ancestors of eukaryotes and eubacteria. The age of the oldest known extinct eukaryotes, *Crypania sipiralis*, was estimated to be about 2 billion years ago, whereas the oldest traces of ancient life on the earth are believed to be approximately 3.6 billion years old [71].

A standard mathematical analysis of the reconstructed phylogenic tree derived from the intron-exon organization of the human *ALDH* genes strongly suggests that *ALDH* genes evolved from an ancestral intronless gene by intron insertion, rather than by a series of intron deletions of an ancestral gene with introns [72].

A very early appearance of ALDH is not surprising, since a non-specific ALDH with a broad substrate specificity was essential to protect early life on the ancient earth which was rich in varieties of highly reactive and toxic aldehydes. A general role of the present day human ALDH isozymes, which display diverse tissue and subcellular distributions, is still detoxification of biogenic and xenobiotic aldehydes existing in given tissues.

ALDHs with more specialized activities had evolved from the ancestral ALDH. ALDH1, ALDH6 and the recently identified ALDH11 which have high activities for oxidation of retinal, had evolved together before the radiation of multicellular organisms.

ALDH3 and ALDH10, which may be involved in the metabolism of membrane lipids, had evolved together during the period of appearance of mammals. Meanwhile, two distantly related enzymes, i.e. ALDH4 which participates in 4-Abu synthesis, and SSDH which participates in 4-Abu degradation, were separated much earlier.

We apologize to many investigators whose work on ALDHs were not cited because of space limitations and because their work is beyond the scope of this review.

Studies in this laboratory were supported by the US Public Health Service (grant HL-29515 to A. Yoshida, GM20293 to M. Nei) and by the National Science Foundation (grant DEB 9520832 to M. Nei).

REFERENCES

1. Racker, E. (1949) Bovine liver aldehyde dehydrogenase, *J. Biol. Chem. 177*, 883–892.
2. Deitrich, R. A. (1966) Tissue and subcellular distribution of mammalian aldehyde-oxidizing capacity, *Biochem. Pharmacol. 15*, 1911–1922.
3. Harada, S., Agarwal, S. & Goodde, H. W. (1980) Electrophoretic and biochemical studies of human aldehyde dehydrogenase isozymes in various tissues, *Life Sci. 26*, 1771–1780.
4. Pietruszko, R. (1983) Aldehyde dehydrogenase isozymes, in *Isozymes: Current topics in biological and medical research*, vol. 8 (Ratazzi, M. C., Scandalios, J. G. & Whitt, A. S., eds) pp. 195–217, Alan R. Liss, New York.
5. Hsu, L. C., Tani, K., Fujiyoshi, T., Kurachi, K. & Yoshida, A. (1985) Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2, *Proc. Natl Acad. Sci. USA 92*, 2703–2707.
6. Chambliss, K. L., Caudle, D. L., Hinson, D. D., Moomaw, C. R., Slaughter, C. A., Jakobs, C. & Gibson, K. M. (1995) Molecular cloning of mature NAD-dependent succinic semialdehyde dehydrogenase from rat and human, *J. Biol. Chem. 70*, 461–467.
7. Kedishvili, N. Y., Popov, K. M., Rougraff, P. M., Zhao, Y., Crabb, D. W. & Harris, R. A. (1992) CoA-dependent methylmalonate semi-aldehyde dehydrogenase, a unique member of the aldehyde dehydrogenase superfamily, *J. Biol. Chem. 267*, 19724–19729.
8. Hsu, L. C., Chang, W.-C. & Yoshida, A. (1989) Genomic structure of the human cytosolic aldehyde dehydrogenase gene, *Genomics 5*, 857–865.
9. Hsu, L. C., Bendel, R. E. & Yoshida, A. (1988) Genomic structure of the human mitochondrial aldehyde dehydrogenase gene, *Genomics 2*, 57–65.
10. Hsu, L. C., Chang, W.-C., Shibuya, A. & Yoshida, A. (1992) Human stomach aldehyde dehydrogenase cDNA and genomic cloning,

primary structure and expression in *E. coli*, *J. Biol. Chem. 267*, 3030–3037.

11. Hsu, L. C., Chang, W.-C., Chang, C., Tsukamoto, N. & Yoshida, A. (1996) The human aldehyde dehydrogenase 3 gene (*ALDH3*): Identification of a new exon and diverse mRNA isoforms, and functional analysis of the promoter, *Gene Expr. 6*, 87–99.
12. Hu, C. A., Lin, W. W. & Valle, D. (1996) Cloning, characterization and expression of cDNAs encoding human Δ^1-pyrroline-5-carboxylate dehydrogenase, *J. Biol. Chem. 271*, 9795–9800.
13. Hsu, L. C. & Chang, W.-C. (1991) Cloning and characterization of a new functional human aldehyde dehydrogenase gene, *J. Biol. Chem. 266*, 12257–12265.
14. Hsu, L. C., Chang, W.-C., Hirnoka, L. & Hsieh, C. L. (1994) Molecular cloning, genomic organization and chromosomal localization of an additional human aldehyde dehydrogenase gene, *ALDH6*, *Genomics 24*, 333–341.
15. Hsu, L. C., Chang, W.-C. & Yoshida, A. (1994) Cloning of a cDNA encoding human ALDH7, a new member of the aldehyde dehydrogenase family, *Gene 151*, 285–289.
16. Hsu, L. C., Chang, W.-C. & Yoshida, A. (1997) Human aldehyde dehydrogenase genes, *ALDH7* and *ALDH8*: Genomic organization and genomic gene structure comparison, *Gene 189*, 89–94.
17. Lin, S. W., Chen, J., Hsu, L. C., Hsieh, C. L. & Yoshida, A. (1996) Human γ-aminobutyraldehyde dehydrogenase (ALDH9): cDNA sequence, genomic organization, polymorphism, chromosomal localization, and tissue expression, *Genomics 34*, 376–380.
18. Chang, C. & Yoshida, A. (1997) Human fatty aldehyde dehydrogenase gene (ALDH10): Organization and tissue-dependent expression, *Genomics 39*, 80–85.
19. Rogers, G. R., Markova, N. G., deLaurenzi, V., Rizzo, W. B. & Compton, J. G. (1997) Genomic organization and expression of human fatty aldehyde dehydrogenase (FALDH), *Genomics 39*, 127–135.
20. Treitel, F., Malaspina, P., Judice, C., Novelletto, A., Slaughter, C. A., Candle, D. L., Hinson, D. D., Chambliss, K. L. & Gibson, K. M. (1997) Human succinic semialdehyde dehydrogenase, *Adv. Exp. Med. Biol. 424*, 253–260.
21. Lee, P., Kuhl, K., Gelbart, T., Kamimura, T., West, C. & Beutler, E. (1994) Homology between a human protein and a protein of green pea, *Genomics 21*, 371–378.
22. Harrington, M. G., Henehan, G. T. M. & Tipton, K. F. (1987) The rat and human aldehyde dehydrogenase isozymes in ethanol metabolism, *Prog. Clin. Biol. Res. 232*, 111–125.
23. Mitchell, D. Y. & Petersen, D. R. (1987) The oxidation of α, β-unsaturated aldehyde products of lipid peroxidation by rat liver aldehyde dehydrogenases, *Toxicol. Appl. Pharmacol. 87*, 403–410.
24. Jakoby, W. B. & Ziegler, D. M. (1990) The enzymes of detoxification, *J. Biol. Chem. 265*, 20715–20718.
25. Yoshida, A., Hsu, L. C. & Davé, V. (1992) Retinal oxidation activity and biological role of human cytosolic aldehyde dehydrogenase, *Enzyme 46*, 239–244.
26. Pereira, F., Roseman, E., Nylen, E., Kaufman, M., Ansky, L. & Wrongemann, K. (1991) The 56 kDa androgen binding protein is an aldehyde dehydrogenase, *Biochem. Biophys. Res. Commun. 175*, 831–838.
27. Yoshida, A., Hsu, L. C. & Yanagawa, Y. (1993) Biological role of human cytosolic aldehyde dehydrogenase-1: Hormonal response, retinal oxidation and implication in testicular feminization, *Adv. Exp. Med. Biol. 328*, 37–44.
28. Zhao, D., McCaffery, P., Ivins, K. J., Neve, R. L., Hogan, P., Chin, W. W. & Drager, U. C. (1996) Molecular identification of a major retinoic-acid-synthesizing enzyme, a retinaldehyde-specific dehydrogenase, *Eur. J. Biochem. 240*, 15–22.
29. Wang, X., Penzes, P. & Napoli, J. L. (1996) Cloning of cDNA encoding an aldehyde dehydrogenase and its expression in *Escherichia coli*, *J. Biol. Chem. 271*, 16288–16293.
30. Yoshida, A. (1983) A possible structural variant of human cytosolic aldehyde dehydrogenase with diminished enzyme activity, *Am. J. Hum. Genet. 35*, 1115–1116.
31. Eckey, R., Agarwal, D. P., Saha, N. & Goedde, H. W. (1986) Detection and partial characterization of a variant cytosolic aldehyde dehydrogenase isozyme, *Hum. Genet. 72*, 95–97.
32. Yoshida, A., Ward, R. J. & Peters, T. J. (1989) Cytosolic aldehyde dehydrogenase (ALDH1) variants found in alcohol flushers, *Ann. Hum. Genet. 53*, 1–7.
33. Hilton, J. (1984) Role of aldehyde dehydrogenase in cyclophosphamide-resistant L1210 leukemia, *Cancer Res. 44*, 5156–5160.
34. Sladek, N. E. & Landkamer, G. L. (1985) Restoration of sensitivity to oxazaphosphorines by inhibitors of aldehyde dehydrogenase activity in cultured oxazaphosphorine-resistant L1210 and cross-linking agent-resistant P388 cell lines, *Cancer Res. 45*, 1549–1555.
35. Yoshida, A., Davé, V., Han, H. & Scanlon, K. J. (1993) Enhanced transcription of the cytosolic *ALDH* gene in cyclophosphamide resistant human carcinoma cells, *Adv. Exp. Med. Biol. 328*, 63–72.
36. Westlund, P., Fylling, A. C., Cederlund, E. & Jörnvall, H. (1994) 11-Hydroxythromboxane B_2 dehydrogenase is identical to cytosolic aldehyde dehydrogenase, *FEBS Lett. 345*, 99–103.
37. King, C. T. & Holmes, R. S. (1997) Human corneal and lens aldehyde dehydrogenases: Purification and properties of human lens ALDH1 and differential expression as major soluable proteins in human lens (ALDH1) and cornea (ALDH3), *Adv. Exp. Med. Biol. 414*, 19–27.
38. Yoshida, A. (1991) Genetic polymorphisms of alcohol metabolizing enzymes related to alcohol sensitivity and alcoholic diseases, *Alcohol Alcohol. 29*, 693–696.
39. Novoradovsky, A. & Goldman, D. (1995) Population genetics of mitochondrial aldehyde dehydrogenase (ALDH2) in American Indian and Asian populations, *Am. J. Hum. Genet. 55*, A966.
40. Shibuya, A., Takeuchi, H., Shibata, K. & Yoshida, A. (1994) Immunohistochemical study of hepatocellular carcinoma-specific aldehyde dehydrogenase, *Alcohol Alcohol. 29*, 119–123.
41. Lindahl, R. (1992) Aldehyde dehydrogenases and their role in carcinogenesis, *Crit. Rev. Biochem. Mol. Biol. 27*, 283–335.
42. Bunting, K. D. & Townsend, A. J. (1996) Protection by transfected rat or human class 3 aldehyde dehydrogenase against the cytotoxic effect of oxazaphosphorine alkylating agents in hamster V79 cell lines, *J. Biol. Chem. 271*, 11891–11896.
43. Sreerama, L. & Sladek, N. E. (1993) Identification and characterization of a novel class 3 aldehyde dehydrogenase overexpressed in a human breast adenocarcinoma cell line exhibiting oxazaphosphorine-specific acquired resistance, *Biochem. Pharmacol. 45*, 2487–2505.
44. Sreerama, L., Rekha, G. K. & Sladek, N. E. (1995) Phenolic antioxidant-induced overexpression of class-3 aldehyde dehydrogenase and oxazaphosphorine-specific resistance, *Biochem. Pharmacol. 49*, 669–675.
45. Sreerama, L. & Sladek, N. E. (1996) Overexpression of glutathione-S transferases, dt-diaphorase and an apparently tumour-specific cytosolic class-3 aldehyde dehydrogenase by Warthin tumours and mucoepidermoid carcinomas of human parotid gland, *Arch. Oral Biol. 41*, 597–605.
46. Yin, S.-J., Cheng, T.-C., Chang, C.-P., Chen, Y.-J., Tang, H. S., Chang, T.-M. & Wu, C. W. (1988) Human stomach aldehyde dehydrogenase (ALDH): A genetic model proposed for ALDH III isozymes, *Biochem. Genet. 26*, 343–360.
47. Tsukamoto, N., Chang, C. & Yoshida, A. (1997) Mutations associated with Sjögren-Larsson syndrome, *Ann. Hum. Genet. 6*, 235–242.
48. Liu, Z.-J., Sun Y.-J., Rose, J., Chung, Y.-J., Hsiao, C.-D., Chang, W.-R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J. & Wang, B.-C. (1997) The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold, *Nat. Struct. Biol. 4*, 317–326.
49. Santisteban, I., Povey, S., West, L. F., Parrington, J. M. & Hopkinson, D. A. (1985) Chromosomal assignment, biochemical and immunological studies on human aldehyde dehydrogenase, ALDH3, *Ann. Hum. Genet. 49*, 87–100.
50. Algar, E. M., Cheung, B., Hayes, J., Holmes, R. S. & Beacham, I. R. (1993) Bovine corneal aldehyde dehydrogenases: Evidence for multiple gene product, *Adv. Exp. Med. Biol. 328*, 153–157.
51. Forte-McRobbie, C. M. & Pietruszko, R. (1986) Purification and characterization of human liver high Km aldehyde dehydrogenase

and its identification as glutamic γ-semialdehyde dehydrogenase, *J. Biol. Chem. 261*, 2154–2163.

52. Efron, M. L. (1965) Familial hyperprolinemia: Report of a second case associated with congenital renal malformations, hereditary and mild mental retardation, with demonstration of an enzyme defect, *New Engl. J. Med. 272*, 1243–1254.
53. Valle, D., Goodman, G. L., Harris, S. C. & Phang, J. M. (1979) Genetic evidence for a common enzyme catalyzing the second step in the degradation of proline and hydroxyproline, *J. Clin. Invest. 64*, 1365–1370.
54. Ambroziak, W. & Pietruszko, R. (1993) Metabolic role of aldehyde dehydrogenase, *Adv. Exp. Med. Biol. 328*, 5–15.
55. Sobue, K. & Nakajima, T. (1978) Changes in concentrations of polyamines and γ-aminobutyric acid and their formation in chick embryo brain during development, *J. Neurochem. 30*, 277–279.
56. Kikonyogo, A. & Pietruszko, R. (1996) Aldehyde dehydrogenase from adult brain that dehydrogenates γ-aminobutyraldehyde: Purification, characterization, cloning and distribution, *Biochem. J. 316*, 317–324.
57. DeLaurenzi, V., Rogers, G. R., Hamrock, J. D., Marekov, L. N., Dompton, J. G., Markova, N. & Rizzo, W. B. (1996) Sjögren-Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene, *Nat. Genet. 12*, 52–57.
58. Jakobs, C., Jacken, J. & Gibson, K. M. (1993) Inherited disorders of GABA metabolism, *J. Inher. Metab. Dis. 16*, 704–715.
59. Yoshida, A. (1992) Molecular genetics of human aldehyde dehydrogenase, *Pharmacogenetics 2*, 139–147.
60. Stewart, M. J., Malek, K., Xiao, Q., Dipple, K. M. & Crabb, D. W. (1995) The novel aldehyde dehydrogenase gene, *ALDH5*, encodes an active aldehyde dehydrogenase enzyme, *Biochem. Biophys. Res. Commun. 211*, 144–151.
61. Goodwin, G. W., Rougraff, P. M., Davis, E. J. & Harris, R. A. (1989) Purification and characterization of methylmalonate-semialdehyde dehydrogenase from rat liver, *J. Biol. Chem. 264*, 14965–14971.
62. Wang, X. & Weiner, H. (1995) Involvement of glutamate 268 in the active site of human liver mitochondrial (class 2) aldehyde dehydrogenase as probed by site-directed mutagenesis, *Biochemistry 34*, 237–243.
63. Steinmetz, C. G., Xie, P., Weiner, H. & Hurley, T. D. (1997) Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion, *Structure 5*, 701–711.
64. Farrés, J., Wang, T. Y., Cunningham, S. J. & Weiner, H. (1995) Investigation of active site cysteine residue of rat mitochondrial aldehyde dehydrogenase by site-directed mutagenesis, *Biochemistry 34*, 2592–2598.
65. Liu Z.-J., Sun Y.-J., Rose J., Chung Y.-J., Hsiao C.-D., Chang W.-R., Kuo I., Perozich J., Lindahl R., Hempel J. & Wang B.-C. (1997) The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold, *Nat. Struct. Biol. 4*, 317–326.
66. Sheikh, S., Ni, L., Hurley, T. D. & Weiner, H. (1997) The potential roles of the conserved amino acids in human liver mitochondrial aldehyde dehydrogenase, *J. Biol. Chem. 272*, 18817–18822.
67. Ni, L., Sheikh, S. & Weiner, H. (1997) Involvement of glutamate 399 and lysine 192 in the mechanism of human liver mitochondrial aldehyde dehydrogenase, *J. Biol. Chem. 272*, 18823–18826.
68. Knoll, A. H. (1991) End of the Proterozoic, *Eur. Sci. Am. 265*, 64–73.
69. Knoll, A. H. (1992) The early evolution of eukaryotes: a geological perspective, *Science 256*, 622–627.
70. Han, T.-M. & Runnegar, B. (1992) Megascopic eukaryotic algae from the 2.1-billion-year-old negaunee iron-formation, Michigan, *Science 257*, 232–235.
71. Schopf, J. W. (1993) Microfossils of the early Archean apex chert: new evidence of the antiquity of life, *Science 260*, 640–646.
72. Rzhetsky, A., Ayala, F. J., Hsu, L. C., Chang, C. & Yoshida, A. (1997) Exon/intron structure of aldehyde dehydrogenase genes supports the "introns-late" theory, *Proc. Natl Acad. Sci. USA 94*, 6820–6825.
73. Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol. 4*, 406–425.
74. Zuckerkandle, E. & Pauling, L. (1965) Evolutionary divergence and convergence in proteins, in *Evolving genes and proteins* (Bryson, V. & Vogel, H. J., eds) pp. 97–166, Plenum Press, New York.
75. Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap, *Evolution 39*, 783–791.
76. Takezaki, N., Rzhetsky, A. & Nei, M. (1995) Phylogenetic test of molecular clock and linearized trees, *Mol. Biol. Evol. 12*, 823–833.
77. Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. (1996) Continental breakup and ordinal diversification of birds and animals, *Nature 381*, 226–229.